

## Data-Mining Technologies for Diabetes: A Systematic Review

Miroslav Marinov, M.S.,<sup>1</sup> Abu Saleh Mohammad Mosa, M.S.,<sup>1</sup> Illhoi Yoo, Ph.D.,<sup>1,2</sup>  
and Suzanne Austin Boren, Ph.D., MHA<sup>1,2</sup>

### Abstract

#### **Background:**

The objective of this study is to conduct a systematic review of applications of data-mining techniques in the field of diabetes research.

#### **Method:**

We searched the MEDLINE database through PubMed. We initially identified 31 articles by the search, and selected 17 articles representing various data-mining methods used for diabetes research. Our main interest was to identify research goals, diabetes types, data sets, data-mining methods, data-mining software and technologies, and outcomes.

#### **Results:**

The applications of data-mining techniques in the selected articles were useful for extracting valuable knowledge and generating new hypothesis for further scientific research/experimentation and improving health care for diabetes patients. The results could be used for both scientific research and real-life practice to improve the quality of health care diabetes patients.

#### **Conclusions:**

Data mining has played an important role in diabetes research. Data mining would be a valuable asset for diabetes researchers because it can unearth hidden knowledge from a huge amount of diabetes-related data. We believe that data mining can significantly help diabetes research and ultimately improve the quality of health care for diabetes patients.

*J Diabetes Sci Technol* 2011;5(6):1549-1556

**Author Affiliations:** <sup>1</sup>Informatics Institute, University of Missouri, Columbia, Missouri; and <sup>2</sup>Department of Health Management and Informatics, University of Missouri School of Medicine, Columbia, Missouri

**Abbreviations:** (CART) classification and regression tree, (BGL) blood glucose level, (FBG) fasting blood glucose, (HbA1c) glycated hemoglobin, (IB1) instance-based learning version 1, (ICU) intensive care unit, (MeSH) medical subject headings, (STS) structural time series, (T1DM) type 1 diabetes mellitus, (T2DM) type 2 diabetes mellitus, (TA) temporal abstraction

**Keywords:** blood glucose level, classification, data mining, diabetes mellitus, feature selection, systematic review

**Corresponding Author:** Illhoi Yoo, Ph.D., CE718 Clinical Support and Education Building, Five Hospital Dr., Columbia, MO 65212; email address [yooil@health.missouri.edu](mailto:yooil@health.missouri.edu)